Roadmap to the Pentium Pro Cluster Workshop

Cluster, cluster on the wall,
which is the fastest of them all?

Armin Mikler
Scalable Computing Laboratory

PPCW, April 10-11, 1997

Outline

- Roadmap of the Workshop:
  - Format
  - Issues
- The SCL Cluster History
- Choices made - Lessons learned
- Ready for low-cost parallel computing

Workshop Format

- Interactive (learn from others' mistakes)
- Panel Discussions
- Lessons learned & Summary Session
- Workshop Results:
  - Cluster Cookbook
  - Workshop Document (on the Web)

Issues

- Hardware
  - motherboard
  - single or multiple processors?
  - how much memory?
  - how much disk space per node?
- Operating System
  - Windows NT
  - Unix
    - Linux
    - FreeBSD
    - NetBSD
  - MVS :-)

Issues cont'd.

- Interconnect
  - Technology
    - 100 Mbps Ethernet
    - Gigabit Ethernet
    - Myrinet
    - others?
  - Bottlenecks?
    - PCI Bus
    - Driver
    - Protocol Stack
  - Topology
    - bus
    - switch
    - ring
    - mesh
    - hypercube
  - Protocols
    - MPI
    - PVM
    - others?
  - Effective Throughput

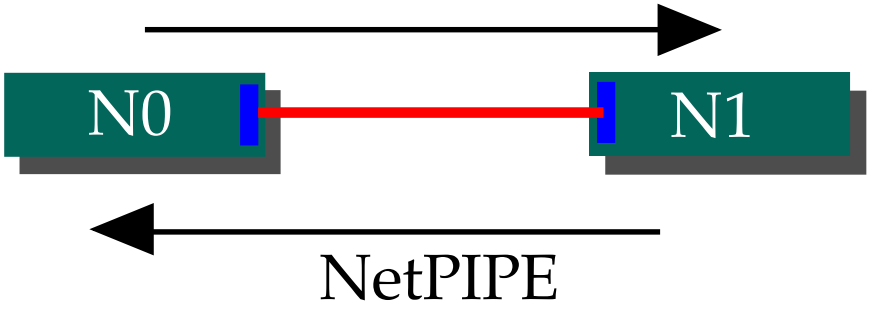
Issues cont'd.

- ♣ Development Tools
  - – Compilers
  - – Debuggers
  - – Performance Monitors
  - – Languages
- ♣ Cluster Configuration
  - – Automated config.
  - – Laundry list for the novice!
- ♣ Administration Tools
  - – User administration
  - – Software management
  - – Scheduling
- ♣ Applications
  - – suitable apps.
  - – cluster limitations

The Beginning

- Tyan Titan Pro S1662 Dual Processor Motherboard with single Processor (200 Mhz), 32 bit, 33 Mhz, 132 Mbps PCI bus with 5 slots, 256KB Secondary Cache, 64 Mb of 60ns EDO, ISA VGA, TEAC Floppy, 2Gb Seagate ST32140A IDE HD
- SMC Etherpower 10/100 (TX version) PCI Fast Ethernet board

Experimental Setups & Upsets

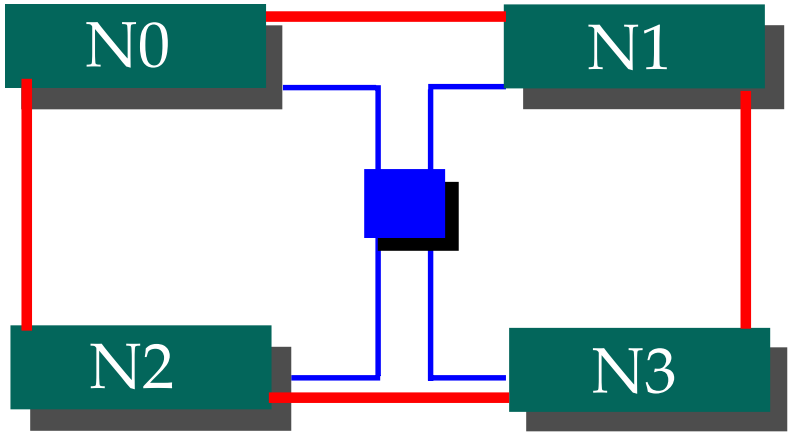


Finding the bottleneck

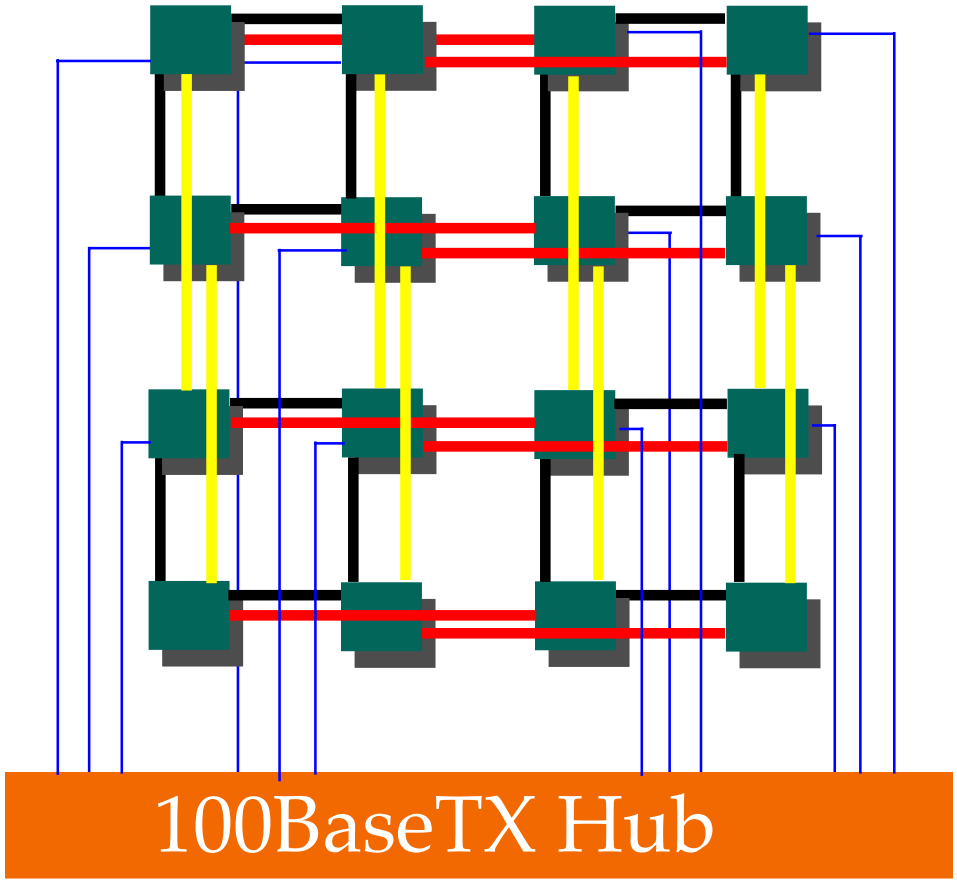
Experiments with multiple networks interfaces.

Issues:

- Node addressing
- Routing

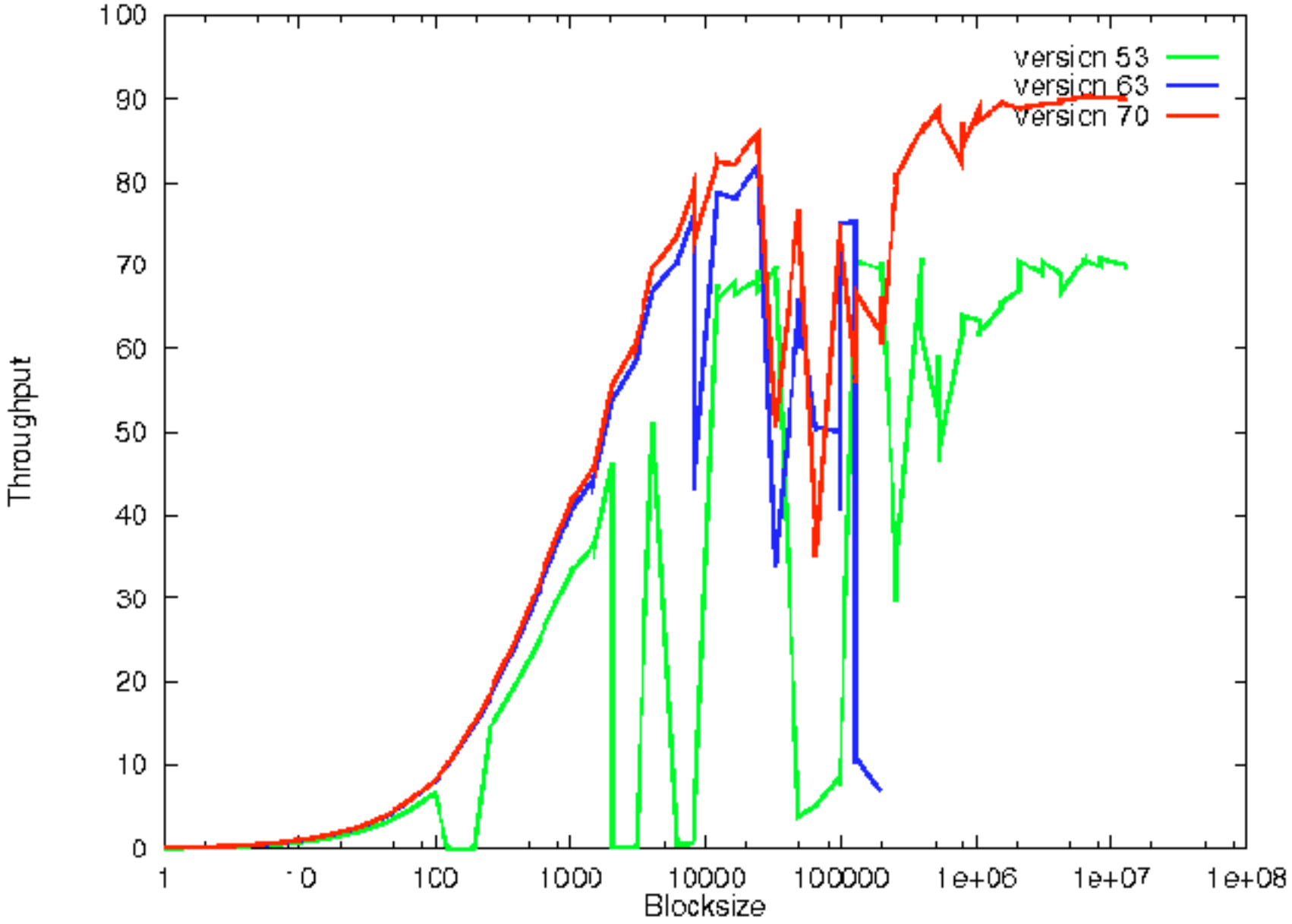


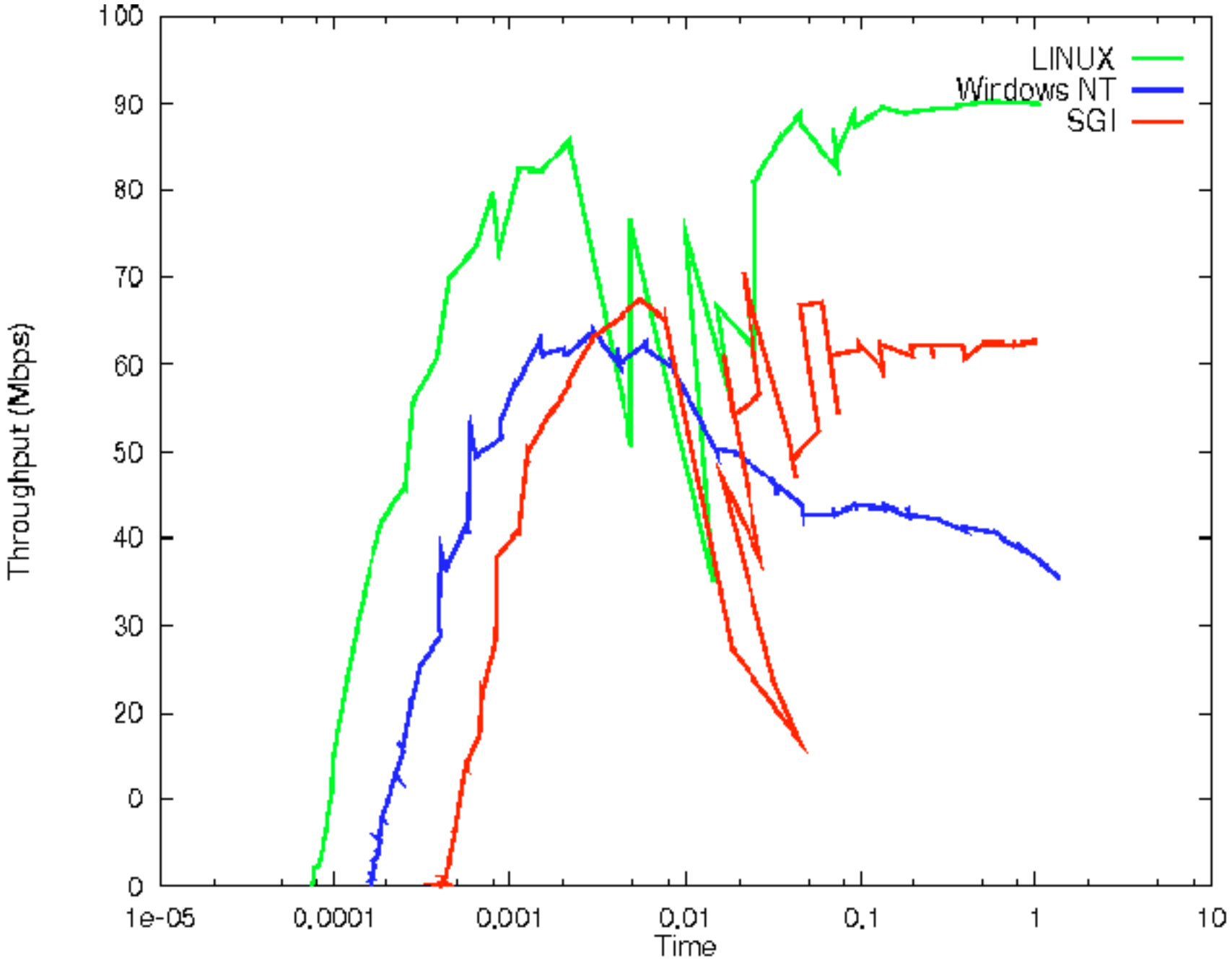
“A Strawman Cluster”

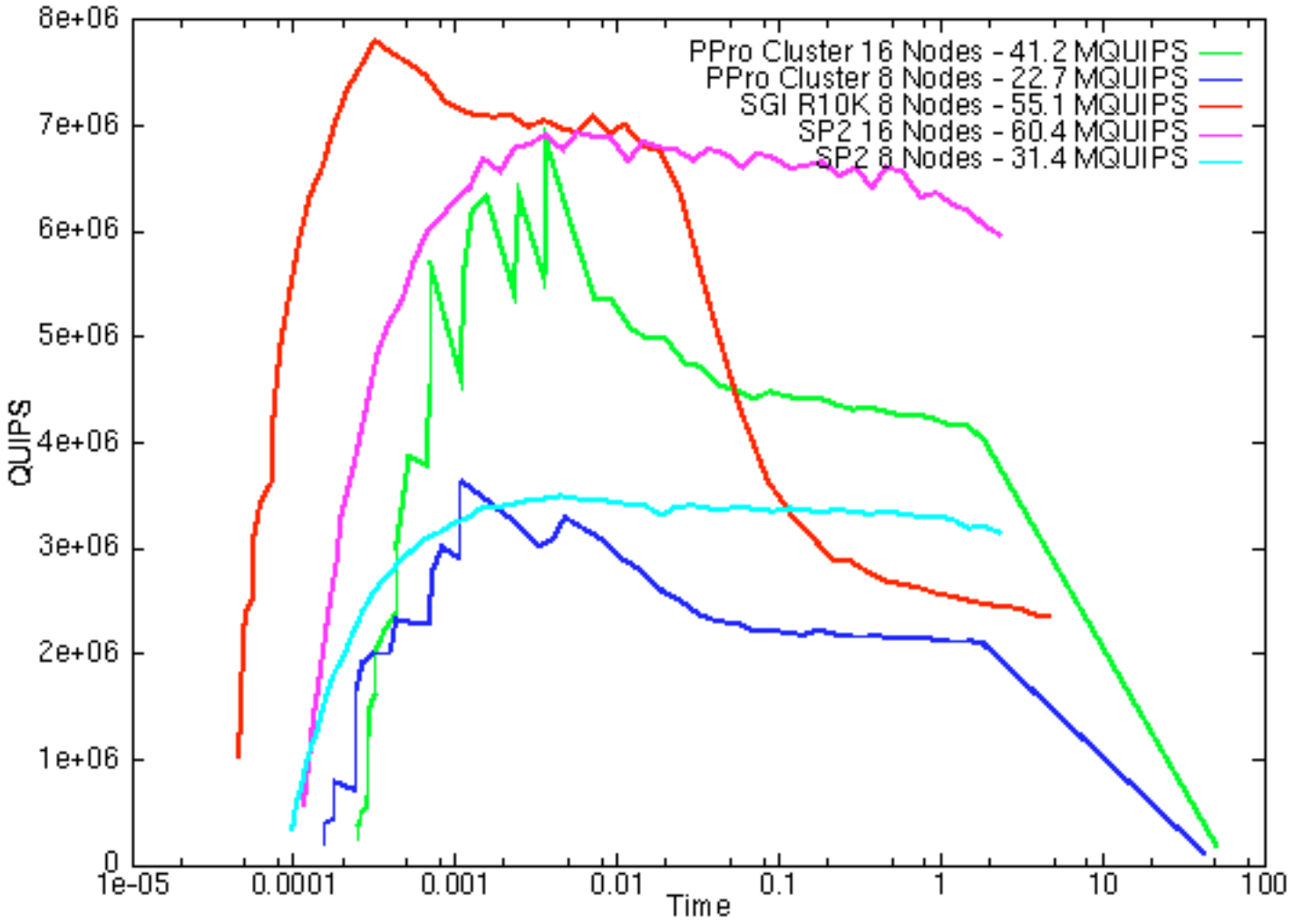


Choices made

- Operating System
  Choices: Windows NT, FreeBSD, Linux
  Rational: Driver support, active user community
- Network Interface
  Choices: SMC single port vs. CogentEM400 TX
  PCI Quartet Fast Ethernet Cards
  Rational: Cost


Throughput
Blocksize
version 53
version 63
version 70
0
10
20
30
40
50
60
70
80
90
100
1
10
100
1000
10000
100000
1e+06
1e+07
1e+08


LINUX
Wirdows NT
SGI
Throughput (Mbps)
Time
100
90
80
70
60
50
40
30
20
0
0
1e-05
0.0001
0.001
0.01
0.1
1
10


PPro Cluster 16 Nodes - 41.2 MQUIPS
PPro Cluster 8 Nodes - 22.7 MQUIPS
SGI R10K 8 Nodes - 55.1 MQUIPS
SP2 16 Nodes - 60.4 MQUIPS
SP2 8 Nodes - 31.4 MQUIPS
QUIPS
Time
8e+06
7e+06
6e+06
5e+06
4e+06
3e+06
2e+06
1e+06
0
1e-05
0.0001
0.001
0.01
0.1
1
10
100